ChinaXiv合作期刊

基于 MODIS 和随机森林的兰州市日最高气温和最低气温估算①

邢立亭， 李 净， 焦文慧
（西北师范大学地理与环境科学学院，甘肃 兰州 730070）

摘 要：近地表气温是衡量城市热环境的重要因素，城市气温被认为是各种城市问题的重要变量，然而小区域气象站严重不足限制了异质城市内气温的空间连续分布表示。为获得空间连续分布的近地表气温，利用遥感数据结合随机森林机器学习法来估算城市近地表空间连续分布的气温。本文以兰州市为研究区，利用 MODIS 传感器的前一天和当天 8 个不同时间点的地表温度数据，结合一系列影响因子，利用随机森林来估算城市的每日最高气温和最低气温（T_{max}/T_{min}）。由于 8 个时序的地表温度数据与 T_{max}/T_{min} 存在不同的相关关系，根据这种关系设计了输入不同地表温度数据的 8 个模型方案，利用实测气温对不同模型方案的结果进行验证，获得最佳方案估算的日最高气温和最低气温。结果表明，用随机森林模型结合遥感数据来估算城市日近地表气温是可行的，并且前一天的地表温度对气温影响较大，是估算气温的关键参数。

关键词：地表温度；MODIS；随机森林；近地表气温；兰州市

chinaXiv:202006.00191v1

城市景观作为地表的一部分，对区域范围内的环境有着巨大的影响，尤其是最高气温和最低气温的估算在人口稠密的地区至关重要。由于城市人口众多且基础设施复杂，城市内气温的微小变化可能会对人类和自然环境产生很大影响[1-4]。

近地表气温是地面气象站测定的常规要素之一，是引起气候变化的关键参数，同时也是研究各种气候模型的一个重要变量。由于城市中气象站点较少且分布分散，利用插值的方法无法准确表示城市气温的空间分布，然而遥感观测能够提供大范围的全球陆面信息，例如地表温度、植被指数等，这使得可以利用遥感获取的与城市气温高度相关的地表信息来估算城市气温[5-8]。

如何获取空间连续分布的近地表气温一直是国内外的研究热点之一。近年来，国内外学者开始利用遥感数据估算近地表气温，但大多以月和年为尺度，很少有以天为尺度的研究[9-12]。目前，比较成熟的气温估算方法可以总结为 4 种[13]：① 简单统计方法，就是根据地表温度和近地表气温的高度相关性，建立简单的一元线性回归模型对气温进行计算，Davis 等[14]根据一元线性回归模型计算得出的北美地区的气温标准差在 1.6～2.6 ℃之间。② 高级统计方法，指在一元线性回归模型的基础上，考虑多个不同因子对气温的影响，如 NDVI、叶面积指数、太阳天顶角、饱和差、相对湿度、经纬度及海拔高度等。常用的 2 个模型为多元回归模型和神经网络模型。Cresswell 等[15]以地温、太阳天顶角为影响因子，建立了与气温的多元回归模型，计算得出的气温值与观测值误差在 0.09～1.69 ℃。③ TVX 法，利用此方法的前提是在气压和湿度条件一致的浓密植被覆盖条件下，植被冠层表面温度接近地表温度且地表温度与 NDVI 呈负相关关系。利用其相关关系来估算气温[16-17]，Stisen 等[18]将 TVX 法结合正弦函数插值法来估算近地表气温，计算得出的均方根误差在 2.55～2.99 ℃，这种方法涉及条件较多且考虑的变量少，气温模拟结果精度较差。④ 能量平衡方法，是指建立地表能量平衡方程来进行气温反演研究，Pape 等[19]结合台站观测数据，利用遥感提取的下垫面植被及地貌数据为参数建立了地表能量平衡方程，得出气温的均方根误差在 0.37～1.02 ℃，由于该方法所需参数较多，空间分布的参数不易获取，使得这种方法在应用上存在一定的限制。

① **收稿日期**：2019-07-01； **修订日期**：2019-09-28
基金项目：国家自然科学基金项目（41761083）；西北师范大学青年教师科研能力提升计划项目（NWNU-LKQN-14-4）资助
作者简介：邢立亭（1996-），男，硕士研究生，研究方向为定量遥感与气温模拟. E-mail：1453374906@qq.com
通讯作者：李净. E-mail：li_jinger@163.com

ChinaXiv合作期刊

在一个城市中自然因素和人为因素在不同空间尺度上混合,使得在城市里估算近地表气温变得难以实现,这使得我们必须探索更为先进的方法来估算气温。近年来,机器学习方法逐渐成熟,越来越多的学者将其应用于不同的领域。在地理学方面,陈浩等[20]利用机器学习中5种模型模拟了青藏高原日降水数据;顾峰等[21]结合Sentinel-2数据,使用随机森林算法估算了干旱区典型绿洲植被叶绿素含量。事实证明机器学习在具有复杂和异质景观的区域中是灵活的,不同学者开始将随机森林方法用于气温的研究[22]。白琳等[23]基于Landsat/TM数据反演了北京市的近地表气温。也有学者使用MODIS地表温度估算近地表气温[24-25],但都是采用当天的瞬时地表温度。由于气温的滞后性,前一天的地表温度与气温的相关性可能会更高,因此,本文以兰州市为研究区,利用8个不同时间的MODIS地表温度产品,并结合高程、NDVI、坡度、坡向和太阳辐射5个因子作为随机森林的输入因子,设计了不同模型方案来估算城市日最高和最低气温,并对不同模型方案的结果进行验证,获得估算城市近地表气温的最佳方案。

1 数据来源与研究方法

1.1 研究区概况与数据来源

1.1.1 研究区概况 兰州市位于甘肃省的中部(图1),其总面积约为13 271 km^2,平均海拔1 500 m,

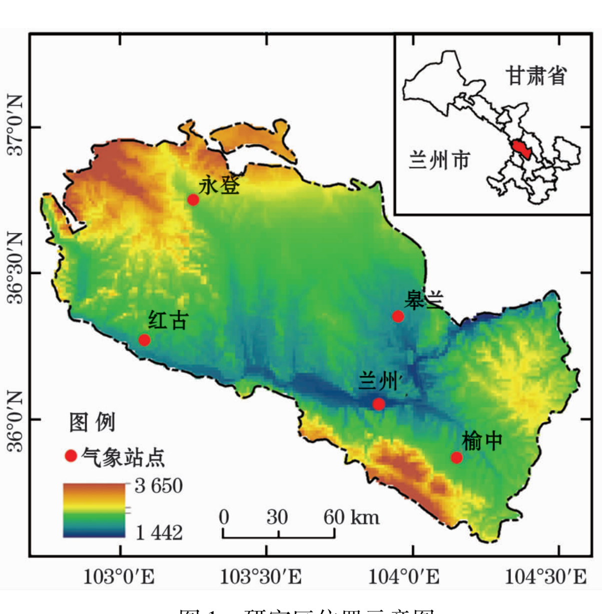


图1 研究区位置示意图

Fig. 1 Location of study area diagram

主体位于南北两山之间的黄河谷地,是典型的河谷城市。属中温带大陆性气候,温差大,降水少,由于主城区位于河谷盆地中,空气的流动性差,城市热岛效应明显。

1.1.2 数据来源 Terra MODIS提供了当地10:30(白天)和22:30(夜间)观测到的地表温度(LST)数据;Aqua MODIS提供了当地时间13:30(白天)和01:30(夜间)观测到的LST数据。本文选取的MODIS数据均来源于NASA官网,其中包括空间分辨率为1 km的地表温度日产品和空间分辨率为1 km的NDVI16天合成产品。研究区行政边界数据来源于国家地球系统科学数据共享平台;DEM数据来源于地理空间数据云平台,空间分辨率为30 m,坡度、坡向数据由DEM获得。太阳辐射数据来自AVHRR卫星获取的日数据,空间分辨率为0.25°,数据来源于https://wui.cmsaf.eu/。气象数据来源于兰州气象局提供的2014年兰州5个站点全年气温日数据。本文用到的数据处理工具主要有MRT、python 2.7.15和ArcGIS 10.4。所用的数据及变量如表1所示。

表1 卫星的衍生变量及研究使用的变量描述

Tab. 1 Description of variables used in this study

变量类型	缩写	简单描述
地表温度(LST)	LST_{AN}	Aqua 卫星当天(01:30,凌晨)的地表温度
	LST_{TD}	Terra 卫星当天(10:30,上午)的地表温度
	LST_{AD}	Aqua 卫星当天(13:30,下午)的地表温度
	LST_{TN}	Terra 卫星当天(22:30,晚上)的地表温度
	LST_{ABN}	Aqua 卫星前一天(01:30,凌晨)的地表温度
	LST_{TBD}	Terra 卫星前一天(10:30,上午)的地表温度
	LST_{ABD}	Aqua 卫星前一天(13:30,下午)的地表温度
	LST_{TBN}	Terra 卫星前一天(22:30,晚上)的地表温度
辅助变量(α)	Sol	每日入射太阳辐射
	NDVI	归一化差异植被指数
	Elev	海拔
	Asp	坡向
	Slo	坡度
气温	T_{max}	最高气温
	T_{min}	最低气温

1.2 研究方法

1.2.1 RF模型 随机森林(random forest)是2001年由LeoBreiman和Culter Adele开发的一种数据挖掘方法[26],与人工神经网络、支持向量机等机器学习方法相比,随机森林算法具有运算量小、容纳样本

chinaXiv:202006.00191v1

ChinaXiv合作期刊

chinaXiv:202006.00191v1

数量大等优点[27]。随机森林方法对非线性数据有着更好的拟合效果,减少了均方根误差、提高了模型的预估精度[28]。

本文通过 Python 2.7.15 中 Pandas 准备数据框数据,再导入 Sklearn 工具包。在 Sklearn 模块库中,利用一系列运算代码实现随机森林模型估算气温。以 8 个地表温度和辅助变量"α"为自变量,因变量为 T_{max} 和 T_{min}。随机森林具体计算步骤如下[29-30]:

(1) 用 Bootstrap 法在 N 个总样本中有放回的随机抽取 n 次,得到 n 个自助样本集作为训练集,未抽取的部分组成袋外数据。

(2) 将每个训练集都单独作为一棵决策树,决策树节点从自变量中选择 M 个(M 小于自变量个数),并按照节点不纯洁度最小原则进行分支生长。

(3) 重复步骤(2) n 次,得到 n 棵决策树组成随机森林。对于每一棵决策树都可以得到一个 OOB 误差估计,将森林中所有决策树的 OOB 误差估计取平均值,可得到随机森林的泛化误差估计。

1.2.2 模型设计 由于本文研究区较小,气象站点的数量较少,为了使模拟时的实测数据更多,模拟结果更加精确,挑选了 2014 年全年 12 d 数据质量较好,且比较有代表性的站点数据(相当于 60 个站点),结合这些天地表温度质量比较好的遥感数据,最终确定了 53 个站点数据,随机选取了 30 个站点,以这些天的数据为例,将这些天不同时间的 8 个地表温度数据分别与城市的 T_{max}(右)和 T_{min}(左)做相关性分析,研究地表温度与最低最高气温的相关关系,结果如图 2 所示。从图 2 可以看出,每天的 T_{max} 和 T_{min} 与不同时间获取的 LST 有不同的关系,Aqua 卫星晚上所获取的地表温度与 T_{min} 的相关性最高,这表明可以用晚上所获取的地表温度数据来估算 T_{min};两个卫星白天所获取的地表温度与 T_{max} 的相关性并不是最高的,Terra 卫星前一天夜晚与 T_{max} 的相关性最高,所以需要设计不同模型,选择最优的方法来估算气温。表 2 分别列出了依据相关分析设计的 T_{max} 和 T_{min} 的不同模型估算方案(S1 ~ S8),S1 方案具有与 T_{max} 或 T_{min} 相关性最高的 LST,S2 方案是在 S1 的基础上添加与 T_{max} 或 T_{min} 相关性第二高的 LST。剩余的方案按相关顺序连续的添加 LST,最后一个方案包含所有的 LST 数据。气温除了与地表温度有关之外,还与 NDVI、太阳辐射、海拔、坡度和坡向有关系,所以,在每种方案中都包含这几个因子,这几个因子用 α 表示(表 1)。设计好方案后,将不同方案输入到随机森林的模型进行估算。

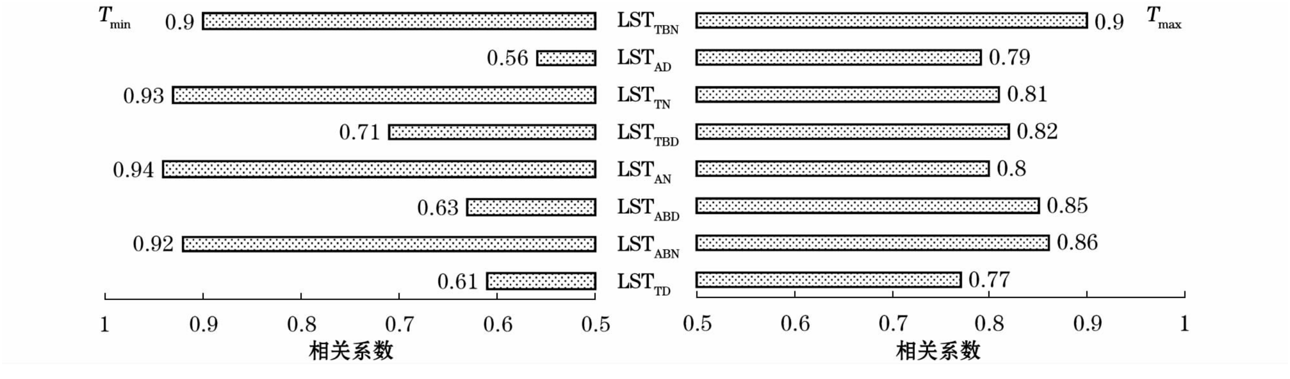


图 2 最高气温和最低气温与地表温度的相关系数

Fig. 2 Coefficient of correlation between T_{max} and T_{min} with surface temperature

2 结果分析

2.1 模型性能评估

在最终确定的 53 个站点数据中选取了 40 个站点作为模型的训练数据,13 个站点(第 97、197 d 和 297 d 站点数据)作为模型的验证数据,模型的验证结果如表 3 所示,模型的 R^2 都超过了 0.8,而设计的这些方案在输入变量数目达到一定的情况下,模型的 R^2 就趋于稳定,甚至 R^2 在达到最高值后会开始降低,随着进入更多的 LST 数据而降低,这是由于数据的冗余所导致的,模型会将多余的数据当做无效数据把其排出在外。在估算最高温度的模型中 S6 显示了最高的性能,模型的 R^2 达到 0.921,估算值与站点实测值之间的 R^2 系数为 0.733,平均绝对误差为 1.344 ℃。平均绝对误差可以准确反映实际估算误差的大小,所以用平均绝对误差来评估方案。而在估算最低气温的模型中,S2 显示了最高的性能,模型的 R^2 为 0.916,估算值与站点实测值之间

ChinaXiv合作期刊

表 2 模型的具体设计方案

Tab. 2 Specific design scheme of model

T_{min}	输入变量	T_{max}	输入变量
S1	$LST_{TBN}+\alpha$	S1	$LST_{AN}+\alpha$
S2	$LST_{TBN}+LST_{ABN}+\alpha$	S2	$LST_{AN}+LST_{TN}+\alpha$
S3	$LST_{TBN}+LST_{ABN}+LST_{ABD}+\alpha$	S3	$LST_{AN}+LST_{TN}+LST_{ABN}+\alpha$
S4	$LST_{TBN}+LST_{ABN}+LST_{ABD}+LST_{TBD}+\alpha$	S4	$LST_{AN}+LST_{TN}+LST_{ABN}+LST_{TBN}+\alpha$
S5	$LST_{TBN}+LST_{ABN}+LST_{ABD}+LST_{TBD}+LST_{TN}+\alpha$	S5	$LST_{AN}+LST_{TN}+LST_{ABN}+LST_{TBN}+LST_{TBD}+\alpha$
S6	$LST_{TBN}+LST_{ABN}+LST_{ABD}+LST_{TBD}+LST_{TN}+LST_{AN}+\alpha$	S6	$LST_{AN}+LST_{TN}+LST_{ABN}+LST_{TBN}+LST_{TBD}+LST_{ABD}+\alpha$
S7	$LST_{TBN}+LST_{ABN}+LST_{ABD}$ D $+LST_{TBD}+LST_{TN}+LST_{AN}+LST_{AD}+\alpha$	S7	$LST_{AN}+LST_{TN}+LST_{ABN}+LST_{TBN}+LST_{TBD}+LST_{ABD}+LST_{TD}+\alpha$
S8	$LST_{TBN}+LST_{ABN}+LST_{ABD}+LST_{TBD}+LST_{TN}+LST_{AN}+LST_{AD}+LST_{TD}+\alpha$	S8	$LST_{AN}+LST_{TN}+LST_{ABN}+LST_{TBN}+LST_{TBD}+LST_{ABD}+LST_{TD}+LST_{AD}+\alpha$

表 3 模型的验证

Tab. 3 Verification of the model

方案(T_{max})	R^2 系数	模型 R^2	MAE/℃	RMSE/℃	方案(T_{min})	模型 R^2	R^2 系数	MAE/℃	RMSE/℃
S1	0.689	0.816	2.332	2.598	S1	0.876	0.784	1.261	1.637
S2	0.691	0.820	2.375	2.650	**S2**	**0.916**	**0.816**	**1.218**	**1.596**
S3	0.697	0.892	1.663	1.801	S3	0.886	0.630	1.305	2.155
S4	0.713	0.912	1.534	1.648	S4	0.882	0.666	1.276	2.049
S5	0.729	0.918	1.513	1.632	S5	0.864	0.686	1.297	2.036
S6	**0.733**	**0.921**	**1.344**	**1.501**	S6	0.870	0.699	1.366	2.056
S7	0.733	0.912	1.352	1.516	S7	0.855	0.701	1.319	2.014
S8	0.706	0.913	1.416	1.538	S8	0.867	0.678	1.361	2.126

注:MAE 为平均绝对误差;RMSE 为均方根误差。加粗的 S2 和 S6 分别为最佳方案。

的 R^2 系数为 0.816,平均绝对误差为 1.218 ℃。

在估算最高气温的模型中,用了前一天地表温度的数据,可见,前一天的地表温度对气温的估算有很大的作用,故可以根据前一天的地表温度来估算后一天的最高气温。在 T_{min} 估算的模型中,前 4 种方案所用到的地表温度数据均为晚上获取的(2 个当天,2 个前天),所以说可以用前一天晚上的地表温度数据来估算后一天的最低气温。统筹这些方案,晚上所获取的地表温度与最高气温和最低气温的相关性都较高。这可能是由于兰州是山谷型城市,白天气温上升不均匀,而晚上在卫星过境时大气的保温作用使气温下降幅度不大的原因。无论是最高气温还是最低气温的估算,随机森林模型都显示了很好的效果,最后绘制了 2014 年第 97、197、297 d 估算的最高和最低气温与站点气温的散点图(图 3)。

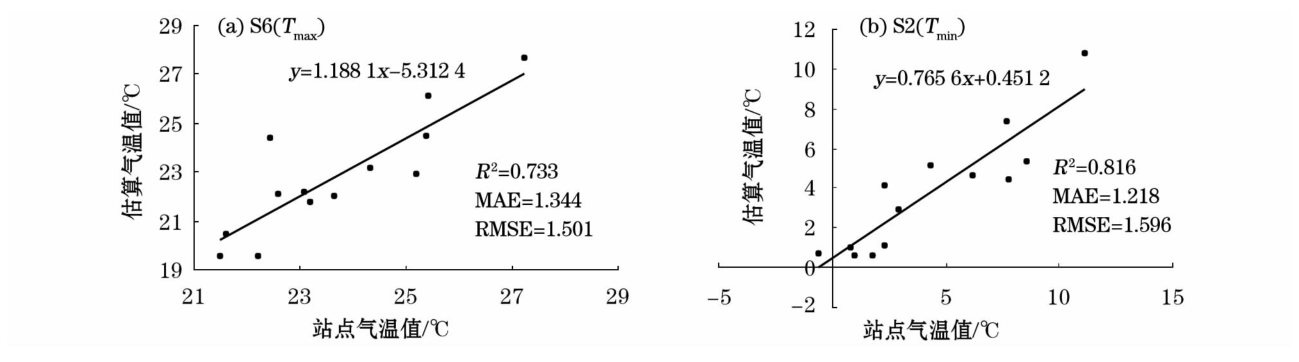


图 3 估算气温与站点气温散点图

Fig. 3 Scatter plot of estimated temperature and site temperature

chinaXiv:202006.00191v1

ChinaXiv合作期刊

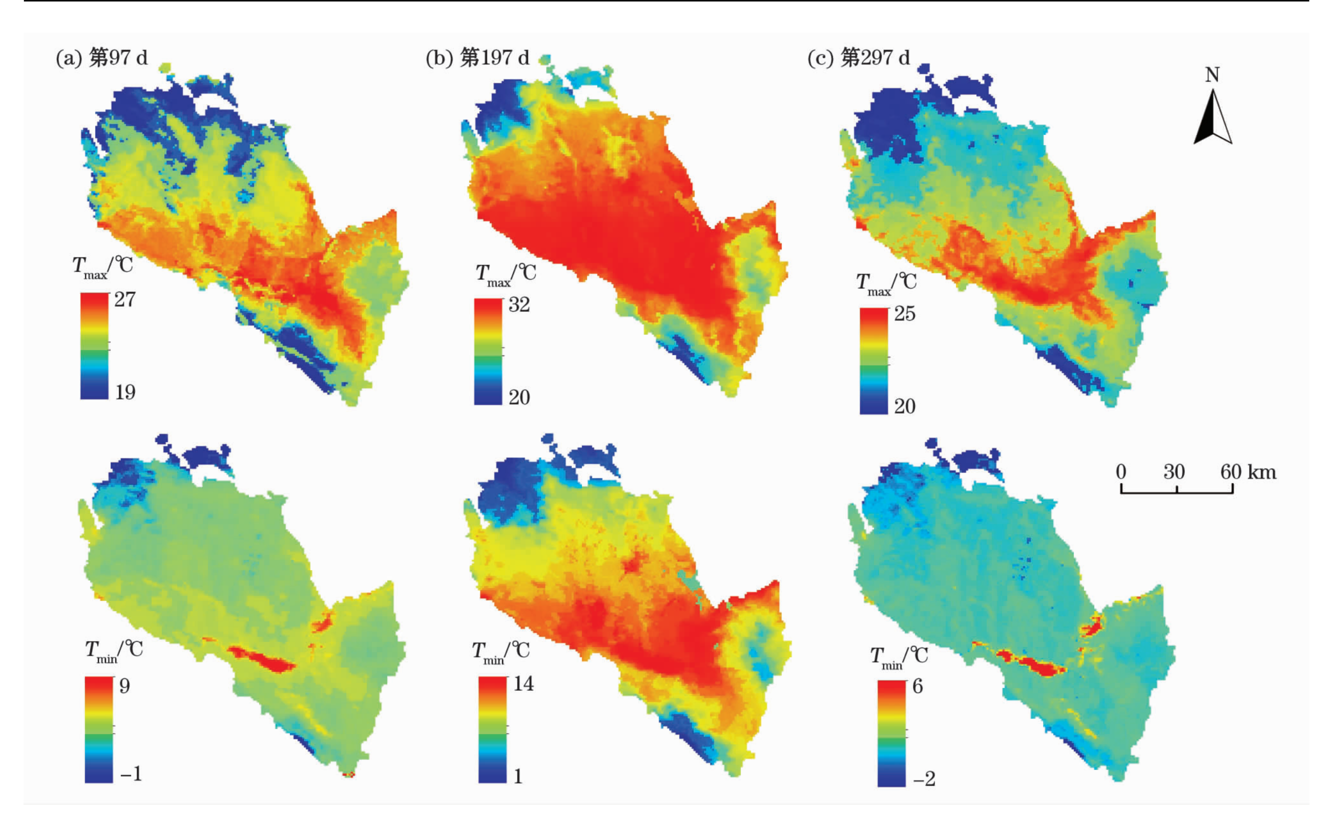


图 4　2014 年第 97、197、297 d 最高/最低气温模拟图

Fig. 4　Maximum/minimum temperature simulation for 97,197,297 days,2014

2.2　空间结果分析

本文以 2014 年第 97、197 d 和 297 d 兰州市数据质量比较好的这几天为例,将地表温度、太阳辐射、海拔、坡度、坡向和 NDVI 数据输入随机森林模型,利用方案 6 和方案 2 分别计算获得了兰州市 2014 年第 97、197 d 和 297 d 近地面的最高气温和最低气温的空间分布图(图 4)。

从图 4 可以看出,用随机森林很好地模拟了气温的空间分布,且每幅图都存在着显著的空间差异性,兰州中心城区和黄河沿岸地区气温较高,呈现出一定的热岛效应。而在南部和北部的山地地区,由于地形和海拔的原因气温比较低。气温呈现出由黄河向南北方向降低的规律,符合海拔越高气温越低的规律。无论是最高气温模型还是最低气温模型所估算的温度都在合理的范围之内,很好地反应了兰州市 2014 年第 97、197 d 及 297 d 最高气温和最低气温的空间分布状况。

3　结论与讨论

本文比较了 8 个不同时间的地表温度和最高、最低气温的相关性,采用随机森林方法结合 MODIS 传感器前一天和当天的多时间序列地表温度数据,估算了兰州市近地表最高和最低气温,评估了 8 种不同的组合方案,获得了估算气温的最佳方案,并用最佳方案模拟了兰州市的最高最低气温。通过本文的研究可得出以下主要结论:

(1) 随机森林结合前一天和当天的 MODIS 地表温度数据,可以很好地估算近地表气温,并且模拟精度较高,是获得空间连续分布气温数据的很好途径。

(2) 利用 MODIS 遥感数据估算气温时,当天所获取的地表温度与当天最高气温及最低气温的相关性并不是最高的,前一日地表温度与最高和最低气温相关性比较高,前一天所获取的地表温度对近地表气温的估算至关重要,在模型设计时必须考虑引入前一天所获取的地表温度。

(3) 在高海拔、地形较为复杂的兰州市,估算气温时除了地表温度之外,还应考虑更多的变量,如高程、坡度、坡向、太阳辐射和归一化植被指数。

本文很好地模拟了不同时间点的连续空间分布气温,具有很好的空间连续性。由于受获取的遥感数据影响,尤其是以日为时间分辨率的研究,时间尺度上不连续,无法揭示出气温的时间变化。所以,在

chinaXiv:202006.00191v1

ChinaXiv合作期刊

chinaXiv:202006.00191v1

利用遥感数据来估算近地表气温的研究还需要不断探索。

参考文献(References):

[1] Yoo C, Im J, Park S, et al. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 137: 149 - 162.

[2] Li L, Zha Y. Estimating monthly average temperature by remote sensing in China[J]. Advances in Space Research, 2019, 63: 2345 - 2357.

[3] 张延伟, 葛全胜, 魏文寿, 等. 应用新方法 HOMR-HOM 均一化 1961—2010 年北疆最高和最低气温[J]. 地理科学, 2015, 35(6): 765 - 772. [Zhang Yanwei, Ge Quansheng, Wei Wenshou, et al. Application of HOMR-HOM for homogenization to North Xinjiang daily maxima and minimum temperature series during 1961 - 2010[J]. Scientia Geographica Sinica, 2015, 35(6): 765 - 772.]

[4] 李国栋, 王乃昂, 张俊华, 等. 兰州市城区夏季热场分布与热岛效应研究[J]. 地理科学, 2008, 28(5): 709 - 714. [Li Guodong, Wang Nai'ang, Zhang Junhua, et al. Urban thermal field and heat island effect of Lanzhou City in summer[J]. Scientia Geographica Sinica, 2008, 28(5): 709 - 714.]

[5] Mao K B, Tang H J, Wang X F, et al. Near-surface air temperature estimation from ASTER data based on neural network algorithm [J]. International Journal of Remote Sensing, 2008, 29(20): 6021 - 6028.

[6] 齐述华, 王军邦, 张庆员, 等. 利用 MODIS 遥感影像获取近地层气温的方法研究[J]. 遥感学报, 2005, 9(5): 570 - 575. [Qi Shuhua, Wang Junbang, Zhang Qingyuan, et al. Study on the estimation of air temperature from MODIS data[J]. Journal of Remote Sensing, 2005, 9(5): 570 - 575.]

[7] 祝善友, 张桂欣. 近地表气温遥感反演研究进展[J]. 地球科学进展, 2011, 26(7): 724 - 730. [Zhu Shanyou, Zhang Guixin. Progress in near surface air temperature retrieved by remote sensing technology[J]. Advances in Earth Science, 2011, 26(7): 724 - 730.]

[8] 韩秀珍, 李三妹, 窦芳丽. 气象卫星遥感地表温度推算近地表气温方法研究[J]. 气象学报, 2012, 70(5): 1107 - 1118. [Han Xiuzhen, Li Sanmei, Dou Fangli. Study of obtaining high resolution near-surface atmosphere temperature by using the land surface temperature from meteorological satellite data[J]. Acta Meteorologica Sinica, 2012, 70(5): 1107 - 1118.]

[9] 任梅芳, 庞博, 徐宗学, 等. 基于随机森林模型的雅鲁藏布江流域气温降尺度研究[J]. 高原气象, 2018, 37(5): 102 - 114. [Ren Meifang, Pang Bo, Xu Zongxue, et al. Downscaling of air temperature in the Yarlung Zangbo River Basin based on the random forest model[J]. Plateau Meteorology, 2018, 37(5): 102 - 114.]

[10] 徐永明, 覃志豪, 沈艳. 基于 MODIS 数据的长江三角洲地区近地表气温遥感反演[J]. 农业工程学报, 2011, 27(9): 63 - 68. [Xu Yongming, Qin Zhihao, Shen Yan. Estimation of near surface air temperature from MODIS data in the Yangtze River Delta[J]. Transactions of the Chinese Society of Agricultural Engineering, 2011, 27(9): 63 - 68.]

[11] Fu P, Weng Q. Variability in annual temperature cycle in the urban areas of the United States as revealed by MODIS imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 146: 65 - 73.

[12] Chen Y, Sun H, Li J. Estimating daily maximum air temperature with MODIS data and a daytime temperature variation model in Beijing urban area[J]. Remote Sensing Letters, 2016, 7(9): 865 - 874.

[13] 张丽文, 黄敬峰, 王秀珍. 气温遥感估算方法研究综述[J]. 自然资源学报, 2014, 29(3): 540 - 552. [Zhang Liwen, Huang Jingfeng, Wang Xiuzhen. A review on air temperature estimation by satellite thermal infrared remote sensing[J]. Journal of Natural Resources, 2014, 29(3): 540 - 552.]

[14] Davis F A, Tarpley J D. Estimation of shelter temperatures from operational satellite sounder data[J]. Journal of Applied Meteorology, 2010, 22(3): 369 - 376.

[15] Cresswell M P. Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model[J]. International Journal of Remote Sensing, 1999, 20(6): 1125 - 1132.

[16] 李斌, 王慧敏, 秦明周, 等. NDVI、NDMI 与地表温度关系的对比研究[J]. 地理科学进展, 2017, 36(5): 585 - 596. [Li Bin, Wang Huimin, Qin Mingzhou, et al. Comparative study on the correlations between NDVI, NDMI and LST[J]. Progress in Geography, 2017, 36(5): 585 - 596.]

[17] 丁海勇, 李往华. 基于 TVX 方法的南京市城区时空格局与地表温度的研究[J]. 长江流域资源与环境, 2018, 27(4): 735 - 744. [Ding Haiyong, Li Wanghua. Analysis of land use land cover temporal-spatial distribution and land surface temperature in Nanjing City using TVX method[J]. Resources and Environment in the Yangtze Basin, 2018, 27(4): 735 - 744.]

[18] Stisen S, Sandholt I, N Rgaard A, et al. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa[J]. Remote Sensing of Environment, 2007, 110(2): 262 - 274.

[19] Pape R, L Ffler J. Modelling spatio-temporal near-surface temperature variation in high mountain landscapes[J]. Ecological Modelling, 2004, 178(3 - 4): 483 - 501.

[20] 陈浩, 宁忱, 南卓铜, 等. 基于机器学习模型的青藏高原日降水数据的订正研究[J]. 冰川冻土, 2017, 39(3): 583 - 592. [Chen Hao, Ning Chen, Nan Zhuotong, et al. Correction of the daily precipitation data over the Tibetan Plateau with machine learning models[J]. Journal of Glaciology and Geocryology, 2017, 39(3): 583 - 592.]

[21] 顾峰, 丁建丽, 葛翔宇, 等. 基于 Sentinel - 2 数据的干旱区典型绿洲植被叶绿素含量估算[J]. 干旱区研究, 2019, 36(4): 924

ChinaXiv合作期刊

－934.［Gu Feng, Ding Jianli, Ge Xiangyu, et al. Estimation of chlorophyll content of typical oasis vegetationin arid area based on Sentinel－2 Data［J］. Arid Zone Research, 2019, 36(4): 924－934.］

［22］华俊玮,祝善友,张桂欣. 基于随机森林算法的地表温度降尺度研究［J］. 国土资源遥感, 2018, 30(1): 78－86.［Hua Junwei, Zhu Shanyou, Zhang Guixin. Downscaling land surface temperature based on random forest algorithm［J］. Remote Sensing for Land & Resources, 2018, 30(1): 78－86.］

［23］白琳,徐永明,何苗,等. 基于随机森林算法的近地表气温遥感反演研究［J］. 地球信息科学学报, 2017, 19(3): 390－397.［Bai Lin, Xu Yongming, He Miao, et al. Remote sensing inversion of near surface air temperature based on random forest［J］. Journal of Geo-Information Science, 2017, 19(3): 390－397.］

［24］毛克彪,马莹,夏浪,等. 用 MODIS 数据反演近地表空气温度的 RM-NN 算法［J］. 高技术通讯, 2013, 23(5): 462－466.［Mao Kebiao, Ma Ying, Xia Lang, et al. An RM-NN algorithm for retrieval of near-surface air temperature retrieval from MODIS data［J］. Chinese High Technology Letters, 2013, 23(5): 462－466.］

［25］姚永慧,张百平. 基于 MODIS 数据的青藏高原气温与增温效应估算［J］. 地理学报, 2013, 68(1): 95－107.［Yao Baihui, Zhang Baiping. MODIS-based estimation of air temperature and heating-up effect of the Tibetan Plateau［J］. Acta Geographica Sinica, 2013, 68(1): 95－107.］

［26］Breiman L. Random forests［J］. Machine Learning, 2001, 45(1): 5－32.

［27］周屹,冯兆祥,白熙卓,等. 基于随机森林算法的数据分析软件设计［J］. 黑龙江工程学院学报, 2017, 31(3): 38－41.［Zhou Yi, Feng Zhaoxiang, Bai Xizhuo, et al. Design of data analysis software based on random forest algorithm［J］. Journal of Heilongjiang Institute of Technology, 2017, 31(3): 38－41.］

［28］刘剑,曹美燕,高治军,等. 一种基于随机森林的太阳能辐射预测模型［J］. 控制工程, 2017, 24(12): 2472－2477.［Liu Jian, Cao Meiyan, Gao Zhijun, et al. A solar radiation prediction model based on radom forest［J］. Control Engineering of China, 2017, 24(12): 2472－2477.］

［29］Breiman L. Bagging preditors［J］. Machine Learning, 1996, 24(2): 123－140.

［30］方匡南,吴见彬,朱建平,等. 随机森林方法研究综述［J］. 统计与信息论坛, 2011, 26(3): 32－38.［Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al. A review of technologies on random forests［J］. Statistics & Information Forum, 2011, 26(3): 32－38.］

Estimation of daily maximum and minimum temperature of Lanzhou City based on MODIS and random forest

XING Li-ting, LI Jing, JIAO Wen-hui

(*College of Geographical and Environmental Science, Northwest Normal University, Lanzhou* 730070, *Gansu, China*)

Abstract: Near-surface temperature is an important factor for measuring the urban thermal environment. Urban temperature is considered to affect various urban problems. However, the meteorological stations in small regions are insufficient to enable sufficient recording of temperature across heterogeneous cities. Remote sensing data combined with the random forest machine learning method was used to estimate the continuous near-surface temperature of Lanzhou. The random forest approach uses the previous day's surface temperature data from a MODIS sensor for eight different time points in the day, combined with a series of influencing factors, to estimate the daily maximum and minimum temperature (T_{max}/T_{min}). Because the eight time points of surface temperature data have different correlations with T_{max}/T_{min}, these different relationships are used to create eight different model schemes with different input surface temperature data, and the results of the different models are verified using the measured temperature to obtain the best estimates of daily maximum and minimum temperatures. We found that it is feasible to use the random forest model combined with remote sensing data to estimate the daily and near surface temperature of Lanzhou city, and that the previous day's surface temperature has a large effect on the present temperature, making it a key parameter for estimating temperature.

Key words: land surface temperature; MODIS; random forest; near surface temperature; Lanzhou City

chinaXiv:202006.00191v1